ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Construct Validation of the Verb Naming Test for Aphasia

Marianne Casilio,[a] Gerasimos Fergadiotis,[b] Sun-Joo Cho,[c] Stacey Steel,[b] Mikala Fleegle,[b] Michael Walsh Dickey,[d] and William Hula[e]

[a]Vanderbilt University Medical Center, Nashville, TN [b]Portland State University, OR [c]Vanderbilt University, Nashville, TN [d]University of Pittsburgh, PA [e]VA Pittsburgh Healthcare System, PA

ARTICLE INFO

Article History:
Received April 8, 2024
Revision received October 7, 2024
Accepted December 19, 2024

Editor-in-Chief: Julie A. Washington
Editor: Amy D. Rodriguez

https://doi.org/10.1044/2024_JSLHR-24-00236

ABSTRACT

Purpose: Although there is widespread agreement pertaining to the cognitive processes underlying spoken word production, more generally in aphasia, multiple competing accounts exist regarding the processes involved for verb production, specifically. Some have speculated that suboptimal control of certain item properties (e.g., imageability) may be partially responsible for conflicting reports in the literature, yet there remains a dearth of research on the psychometric validation of verb production tests for aphasia. The purpose of the present study was to investigate the cognitive constructs underlying the Verb Naming Test (VNT), a relatively commonly used verb production test, by expanding upon an item response theory (IRT) modeling framework we previously described.
Method: Using an archival data set of 107 individuals with aphasia, we specified a series of IRT models to investigate whether item covariates (argument structure, imageability), person covariates (aphasia subtype, severity), and their interactions were predictive of VNT item response patterns.
Results: Across all models, covariates that were most strongly associated with lexical-semantic processing (imageability, aphasia severity) were significant predictors. In contrast, covariates that were most strongly associated with morphosyntactic processing (argument structure, aphasia subtype) were minimally predictive.
Conclusions: VNT item response patterns appear to be primarily explained by covariates representing lexical-semantic processing. In particular, we identified an important role of imageability, a covariate not controlled for in the VNT's item design, which both aligns with a body of prior research and further illustrates the challenge of differentiating morphosyntactic processing from lexical and semantic processes during word production.
Supplemental Material: https://doi.org/10.23641/asha.28664669

Spoken word production is understood to be a rapid and integrative system involving multiple processes (e.g., Dell et al., 1997; Foygel & Dell, 2000; Levelt et al., 1999; Walker & Hickok, 2016). Despite their differences, most accounts agree on the following: In order to produce a word, conceptual representations are activated and mapped to a lemma or word representations (lexical-semantic processing) that is assigned grammatical markers (morphosyntactic processing) and various sound representations (phonological processing) and then motorically planned and executed (speech motor processing). Taken collectively, these processes are thought to represent the core components of word production.

Disruptions to word production are commonly referred to as anomia, a hallmark characteristic of aphasia (Benson, 1979; Goodglass & Wingfield, 1997) and one that is highly associated with overall aphasia severity (e.g., Kristinsson et al., 2023; Walker & Schwartz, 2012; Wilson et al., 2023). Response analyses of the types of errors (e.g., semantic vs. phonemic) produced during confrontation naming in people with aphasia suggest that anomia predominantly reflects lexical-semantic and phonological

Correspondence to Marianne Casilio: marianne.e.casilio@vanderbilt.edu. *Disclosure:* *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

processes (e.g., Dell et al., 2004; Mitchum et al., 1990; Schwartz et al., 2006), although speech motor processing also may play a substantive role (e.g., Walker & Hickok, 2016; Walker et al., 2018). Notably, such studies emphasize the production of nouns in isolation, which minimally tax morphosyntactic processing and purposefully manipulate properties understood to involve lexical-semantic and phonological processing, such as age of acquisition and word length (e.g., Kittredge et al., 2008; Whitworth et al., 2013).

By contrast, agrammatism is also a commonly observed symptom in aphasia (e.g., Goodglass, 1997; Kertesz, 1979; Wilson et al., 2023) that is putatively associated with the Broca's aphasia syndrome (Benson, 1967). Agrammatism is characterized by omissive errors of inflectional morphemes and closed class words (e.g., articles, prepositions) that are thought to reflect a selective disruption to the morphosyntactic component of word production. However, there remains substantial debate as to whether agrammatism instead reflects strategic compensation (e.g., Fedorenko et al., 2022; Kolk & Heeschen, 1992; Rezaii et al., 2022, 2023) that may not be clearly dissociated from concomitant disruptions of speech motor processing (e.g., Benson, 1967; Casilio et al., 2025; Lorca-Puls et al., 2023). In studies of agrammatism, behaviors are commonly identified in elicited sentence production tasks, where sentence properties are manipulated on the basis of syntactic complexity, such as passive sentence structures, embedded clauses, or complex inflectional morphology (Faroqi-Shah & Dickey, 2009; Faroqi-Shah & Thompson, 2007; Friedmann & Grodzinsky, 1997; Wilson et al., 2014). Such behaviors have also been identified in metrics derived from connected speech (e.g., Casilio et al., 2019; Malyutina et al., 2016; Saffran et al., 1989; Vermeulen et al., 1989).

The cognitive processes involved in the production of verbs have been hypothesized to lie at the nexus of those important to both anomia (i.e., lexical-semantics) and agrammatism (i.e., morphosyntax; Chang et al., 2006; Gordon & Dell, 2003), and many argue these processes can be productively captured during confrontation naming (e.g., Kim & Thompson, 2000; Zingeser & Berndt, 1990). As a result, verb production impairments in aphasia have received considerable attention, with multiple competing accounts on the cognitive processes that underlie them. In one view, studies revealing dissociations in verb and noun production among participants classified with anomic and agrammatic aphasia have led scholars to hypothesize an underlying lexical deficit, where word forms have separate stores that are specific to grammatical class (e.g., Caramazza & Hillis, 1991; Miceli et al., 1984, 1988). Others have argued that any dissociations or relative divergences in verb and noun production instead reflect a semantic deficit, as verbs tend to possess fewer perceptual associations than nouns (E. Bates et al., 1991; Bird, Lambdon, et al., 2003; Bird et al., 2000; Luzzatti et al., 2002; Vigliocco et al., 2011). From another viewpoint, these dissociations indicate that verb production is driven by morphosyntactic processing, a perspective informed by the observation of worse performance on verbs with a greater number of arguments or inflectional morphemes in people classified with agrammatic aphasia as compared to those with anomic aphasia (Faroqi-Shah & Thompson, 2007; Friedmann & Grodzinsky, 1997; Kim & Thompson, 2000; Saffran et al., 1980; Thompson, 2003; Thompson et al., 2012). A fourth and final viewpoint argues that verb production impairments are fundamentally multifactorial, involving differential involvement of any and all core processes, with no clear dissociations relative to noun production or in particular aphasia subtypes such as anomic or agrammatic (e.g., Alyahya et al., 2018; Basso et al., 1990; Black & Chiat, 2003). Notably, such studies have argued that suboptimal control of the properties of the test items (e.g., imageability, or how readily a mental image can be generated in response to a stimulus) used to measure verb production, along with the reliance on case studies or small samples, has inflated the observation of dissociations that may instead be artifacts of study design or rare in occurrence. Although rare cases are unquestionably of value in elucidating the underlying architecture of cognitive processing, such reports are not immune to methodological limitations, and findings from them should ultimately be corroborated relative to other case studies or group studies (Lambon Ralph et al., 2011; Schwartz & Dell, 2010).

Although noun production tests of confrontation naming have undergone extensive validation (Fergadiotis et al., 2015, 2019, 2021; Fergadiotis, Swiderski, et al., 2019; Hula et al., 2015, 2020), there remains minimal understanding of the psychometric properties of analogous verb production tests. Thus, the degree to which spurious aspects of test or study design may be contributing to these conflicting accounts of verb production in aphasia is ultimately unclear. Only one relatively common verb production test for aphasia, the Verb Naming Test (VNT; Cho-Reyes & Thompson, 2012), has emerging evidence regarding its psychometric properties. A subtest of the Northwestern Assessment of Verbs and Sentences (NAVS; Cho-Reyes & Thompson, 2012), the VNT consists of 22 items and requires that individuals produce a target verb in response to a pictorial scene (i.e., confrontation naming). It was designed to differentially tax verb-related morphosyntactic processing while holding lexical-semantic and phonological processing relatively constant. Specifically, lexical frequency and phoneme (sound) length of the targeted verbs is controlled for in the VNT. In contrast, targeted verbs vary in their argument structure—the number of participant roles or object phrases associated with the verb phrases containing the target verb. It is worth noting that argument structure is ultimately relevant to both lexical-semantic and morphosyntactic processing in word and sentence production. Specifically, the number of arguments and their semantic relationship to the verb are part of a verb's lexical-semantic representation (Dowty, 1991; Ferretti et al., 2001; Levin, 1993; Mauner, 2015), and there is robust evidence of semantic priming between verbs and arguments (e.g., McRae et al., 2005). Despite this overlap, argument structure is relatively easy to measure and manipulate in a confrontation naming context as compared to other possible properties (e.g., inflectional morphology), thus suggesting it is a practical proxy for verb-related morphosyntactic processing.

Under a classical test theory (CTT) framework, the VNT has been shown to possess adequate interrater reliability and is discriminative of persons with agrammatic and anomic aphasia, as assessed in 59 participants classified with the two subtypes (Cho-Reyes & Thompson, 2012). More recently, we evaluated the psychometric properties of the VNT (Fergadiotis et al., 2023). Here, we used item response theory (IRT), a modern approach to psychometric validation that, in its simplest form, operationalizes the probability of a correct response as a function both of a person's underlying ability and the difficulty of a given item (Lord, 1980). Using data from 107 participants with aphasia with diverse impairment profiles, we found the VNT demonstrated adequate fit to a one-parameter logistic (1-PL) model and possessed good reliability under both CTT and IRT frameworks. Importantly, through these analyses, we found that responses to items on the VNT demonstrate an adequate unidimensional structure[1] per results from both exploratory and confirmatory factor analyses (see Fergadiotis et al., 2023, for complete details). However, the constructs that influence that single dimension have yet to be explored. In other words, it remains an open question whether verb production ability as measured by the VNT indeed is determined primarily by verb-related morphosyntactic processing, as the developers hypothesized, or whether other cognitive processes have greater influence on performance. Specifically, does verb production ability, as measured by the VNT, instead primarily reflect lexical and/or semantic processing, in line with other competing accounts of verb production impairments?

The purpose of the present study was to extend our prior work (Fergadiotis et al., 2023) on the VNT by investigating its construct validity using IRT. Specifically, we aimed to determine whether item and person covariates that are associated with two core constructs, or processes, of verb production—lexical-semantic processing versus morphosyntactic processing—were predictive of different VNT item response patterns. We had three research questions: (1) Does VNT item difficulty relate more to item properties that are assumed to be representative of morphosyntactic or lexical-semantic processing? (2) Does verb production ability (i.e., latent test-level score) dissociate in those with and without agrammatic aphasia? (3) Does the probability of a correct response on the VNT depend on the interaction between these item and person covariates?

[1]Importantly, we do not claim that verb production is a unidimensional cognitive process, rather that accuracy scores on the VNT can be usefully represented by a unidimensional measurement model. In this context, unidimensionality can be understood as the claim that a set of test items all respond to a common underlying factor, with no substantial structure in the residual covariances among the items. Put another way, item responses on the VNT can be understood as psychometrically unidimensional, although verb production may be psychologically multidimensional (Henning, 1992). The explanatory IRT model described below seeks to understand the influences on the hypothesized underlying factor/dimension. An alternative approach to evaluating multidimensionality using the VNT, both psychologically and psychometrically, is to contrast unidimensional and multidimensional IRT models (e.g., Horton et al., 2012). Although not the approach taken in the present study, given that VNT item responses demonstrated adequate model fit to a unidimensional measurement model (Fergadiotis et al., 2023), it nonetheless could be a useful and productive direction for future research on the construct validity of VNT and other verb production tests in aphasia.

Method

Participants

As described by Fergadiotis et al. (2023), audiovisual recordings of VNT administrations from 107 participants with aphasia across 20 data collection sites were obtained from an archival data set (AphasiaBank; MacWhinney et al., 2011). Right-handed native English speakers with a diagnosis of aphasia following a single left-hemisphere stroke who had adequate hearing and vision (aided or unaided) were included. Aphasia diagnosis was operationalized as an Aphasia Quotient (AQ) on the Western Aphasia Battery–Revised (WAB-R; Kertesz, 2006) of < 93.8 or an overall score on the short form of the Boston Naming Test (Kaplan et al., 2001) of < 11. Those with major psychiatric or neurologic disorders (e.g., dementia) were excluded. Demographic and clinical information are presented in Table 1. Given the use of an extant data set, the present study was exempt from approval by an institutional review board.

VNT Item Responses

Dichotomous scores (i.e., 1 = *correct*, 0 = *incorrect*) for responses from all participants were generated in the

Table 1. Demographic and clinical characteristics of the participant sample.

Characteristic	Value
Age (years)	
M (*SD*), range	60 (10.4), 39.5–85.7
Education (years)	
M (*SD*), range	14.7 (2.3), 11–20
Missing (%)	3.7
Ethnicity (%)	
African American	14
Asian	0.9
White	84.1
Gender (%)	
Female	39.8
Male	40.2
Poststroke onset (years)	
M (*SD*), range	5.1 (4.1), 0.3–24.7
Data collection site (# participants)	
M (*SD*), range	5.9 (3.5), 1–13
WAB-R Aphasia Quotient (1–100)	70 (17.1), 20.5–97.9
BNT, Short Form (% correct)	
M (*SD*), range	46.5 (29.2), 0–100
VNT (% correct)	
M (*SD*), range	65.6 (29.4), 0–100
Apraxia of speech (%)	
Present	45.8
Not present	46.7
Unknown	7.5
Dysarthria (%)	
Present	9.3
Not present	80.4
Unknown	10.3

Note. WAB-R = Western Aphasia Battery–Revised (Kertesz, 2006); BNT = Boston Naming Test (Kaplan et al., 2001); VNT = Verb Naming Test (Cho-Reyes & Thompson, 2012).

same manner as our prior work (Fergadiotis et al., 2023; see the Method section and the Appendix for complete details). In brief, extracted videos were reviewed, and responses were transcribed independently by two trained research assistants. Discrepancies between the two research assistants were resolved by a licensed speech-language pathologist. Consensus transcripts were then scored in the same manner following published criteria (Cho-Reyes & Thompson, 2012) with some minor adjustments (Fergadiotis et al., 2023). Following the VNT scoring manual, responses were considered correct if they contained the targeted verb, regardless of whether inflectional morphology was absent or mis-selected. Semantically similar verbs containing the same argument structure as the targeted verb (e.g., saying "giggle" for the item *laugh*) were also permitted, as were pronunciation errors, either phonemic or articulatory in nature, that did not obscure identification of the response verb and affected < 50% of its phonemes (see Cho-Reyes & Thompson, 2012, for additional details regarding VNT scoring procedures). Scored item-level responses are provided in Supplemental Material S1.

Covariates

All relevant person and item covariates, as described below, were fully standardized (i.e., *z* score transformed) prior to analysis to place all variables on a common scale and thus simplify comparisons across variables when interpreting the results. These are included alongside the scored item-level responses in Supplemental Material S1.

Item Covariates

As described in detail below, we focused our main analyses on two primary item covariates—argument structure and imageability. Three other item covariates representative of lexical-semantic and phonological processing (lexical frequency, age of acquisition, phoneme length) and known to be predictive of item difficulty on a common test of confrontation naming of noun production (Fergadiotis, Swiderski, et al., 2019) were also evaluated in a preliminary analysis (see Supplemental Material S4 for a description of these covariates). This analysis was done to confirm that these covariates were adequately controlled for in the VNT item design, as intended by its developers (Cho-Reyes & Thompson, 2012).

Argument structure. As briefly outlined in the first part, items on the VNT are manipulated on the basis of the number of optional or obligatory arguments within the predicate of an utterance (Cho-Reyes & Thompson, 2012), as obtained from the Brandeis Verb Lexicon (Grimshaw & Jackendoff, 1981). For the VNT, the number of verb arguments ranges from 1 to 3 (M = 2.09; SD = 0.75), with a greater number of arguments resulting in a verb that is more difficult to accurately produce (Cho-Reyes & Thompson, 2012). This sensitivity of verb argument structure has been previously documented in studies of people with agrammatic aphasia (Cho-Reyes & Thompson, 2012; De Bleser & Kauschke, 2003; Kim & Thompson, 2000; Kiss, 1997; Thompson et al., 1995, 1997), thus further suggesting that argument structure may be indexing verb-related morphosyntactic processing. As such, for the purposes of the present study, the number of arguments for each of the VNT's 22 items, defined during test design (Cho-Reyes & Thompson, 2012), was extracted as an item covariate. Notably, we did not distinguish between optional and obligatory argument structures, as this distinction has previously been found to not be sensitive to the presence of agrammatism (Cho-Reyes & Thompson, 2012), suggesting it may not be specific to verb-related morphosyntactic processing.

Imageability. In contrast, imageability was not explicitly measured as part of the VNT's test design (Cho-Reyes & Thompson, 2012). This item covariate, which captures the degree to which a given word evokes a mental image, has previously been shown to be sensitive to test-level performance on confrontation naming of verb production (Alyahya et al., 2018; Basso et al., 1990; Luzzatti et al., 2002), where less imageable items were more difficult to accurately produce. Imageability has been widely presumed to be a proxy of semantic-conceptual processing (Nickels et al., 2022), although distinguishing conceptual-semantic from lexical processing is not straightforward (e.g., Franklin et al., 1995; Johnson et al., 1996; Lampe et al., 2021). Thus, imageability here is interpreted as broadly capturing the activation of semantic concepts and their mapping to lexical forms. For the present study, we quantified imageability by extracting averaged 1–7 ratings of imageability, where higher values connote a greater ease in conjuring a mental image of the verb, for each of the 22 VNT items (M = 4.81; SD = 0.80) from the South Carolina Psycholinguistic Metabase (Gao et al., 2022). Specifically, we used ratings from Graves et al. (2010), which are a compilation of ratings from six sources (Bird et al., 2001; Clark & Paivio, 2004; Cortese & Fugett, 2004; Gilhooly & Logie, 1980; Paivio et al., 1968; Toglia & Battig, 1978).

Person Covariates

As described in detail below, two person covariates—WAB-R subtype and AQ—were selected in order to replicate and extend findings important to the original psychometric validation of the VNT (Cho-Reyes & Thompson, 2012).

WAB-R subtype. The prior finding of worse performance on the VNT for those with agrammatic aphasia was based primarily on group classifications using WAB-R subtypes (Cho-Reyes & Thompson, 2012). Specifically, agrammatic aphasia was defined as those with the Broca's subtype, and anomic aphasia was defined as those with the Anomic subtype. Thus, for the present study, subtype classifications were extracted from behavioral testing available as part of AphasiaBank (MacWhinney et al., 2011). We retained the Broca's (n = 30) and Anomic (n = 37) subtypes as specified in the WAB-R; the remaining subtypes within our data set (Global, Wernicke, Conduction, Transcortical Motor, Not Aphasic) were classified collectively as "Other" (n = 40). Subtypes were then specified as a three-level, deviance-coded factor with Broca's as the reference category. Deviance coding was used to express Item × Person interactions in a manner similar to analysis of variance (ANOVA) and to make the unit of analysis for subtype, a categorical variable, comparable to a continuous scale.

WAB-R AQ. AQ (0–100, where higher scores indicate relatively less impairment), a marker of overall aphasia severity, was additionally included as a person covariate for two reasons. First, prior observations have shown the Broca's subtype is associated with more severe impairment overall than the Anomic subtype (e.g., Kertesz, 2006). Second, a preliminary review of AQ scores among the Broca's, Anomic, and Other subtypes within our data set showed significant mean differences as per a one-way ANOVA with equal variances, $F(2, 104) = 52.56$, $p < .001$. Thus, to control for any confounding effect of severity on meaningful differences in verb production performance across the aphasia subtypes of interest, AQ was added as a controlling covariate (see Table 1 for descriptive statistics). As with the WAB-R subtype covariate, AQ was extracted for all participants from AphasiaBank (MacWhinney et al., 2011).

Item × Person Covariate Interactions

Item × Person interaction terms were additionally specified, as we hypothesized that the influence of item covariates on VNT item response patterns may vary as a function of person covariates. Specifically, we were interested in whether the effect of argument structure or imageability moderated the effect of a person's WAB-R subtype or AQ on VNT item response patterns.

Analyses

To address our three research questions, we specified multiple 1-PL IRT models of progressive complexity using the "glmer" function of the *lme4* package (D. Bates et al., 2015) in R. Models were run using Laplace approximation, an efficient algorithm for marginal maximum likelihood estimation. Prior to performing our analyses, a small number of VNT item responses that were missing at random (see Fergadiotis et al., 2023, for details) were item-level deleted (i.e., only a given item response for a particular person was deleted and all other data were retained).

General IRT Modeling Framework

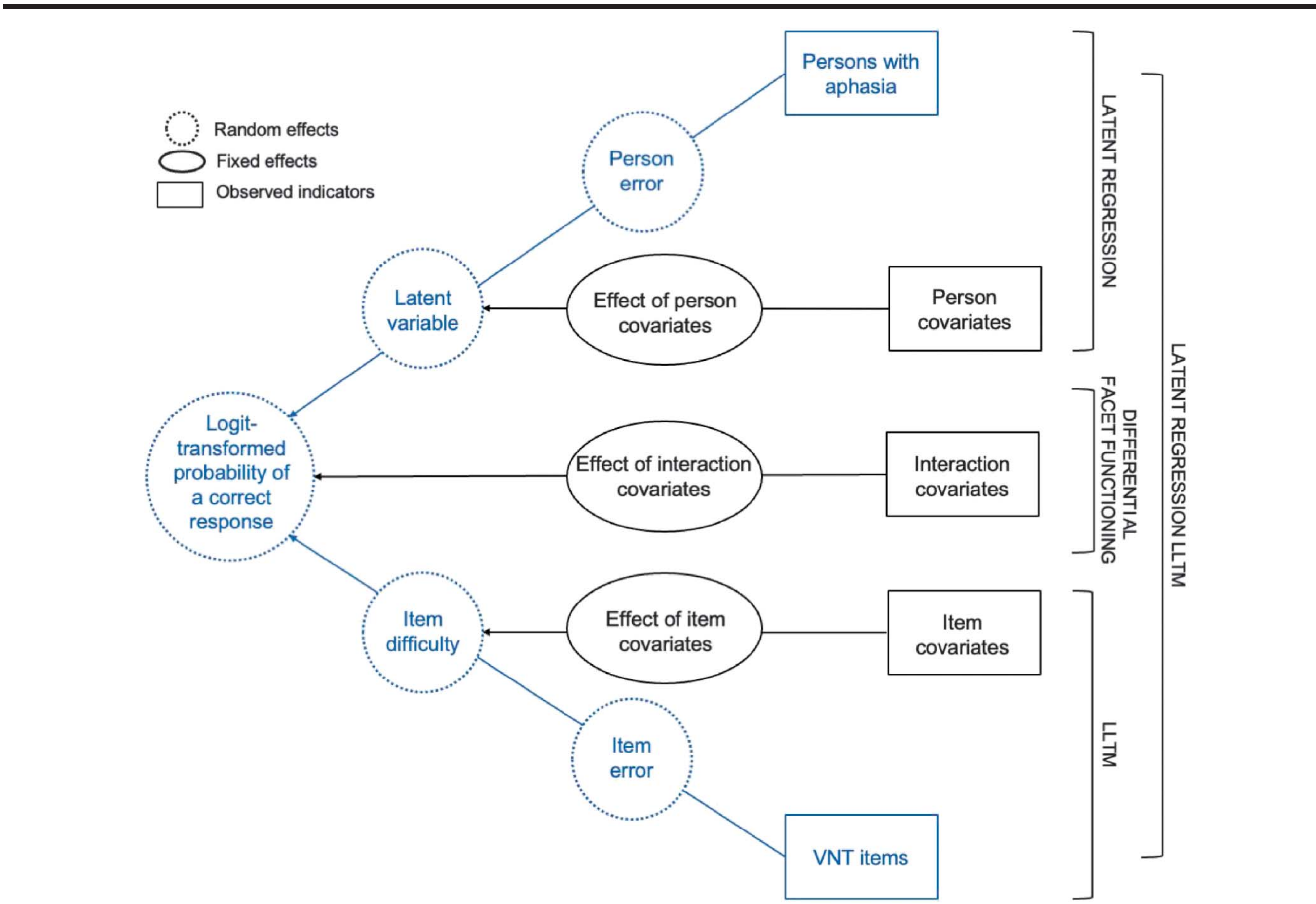
The foundation of the modeling framework for the present study is the 1-PL IRT model, which we previously identified as an appropriate measurement model[2] for the VNT (Fergadiotis et al., 2023). The 1-PL is the simplest

[2]Beyond demonstrating a unidimensional structure, as mentioned in the first part, all 1-PL IRT models additionally must meet the assumptions of local independence, where responses to each item are independent from one another after accounting for all items collectively measuring a single dimension, and equal discrimination, where items are shown to be invariant in their discrimination of individuals across levels of latent ability (Lord, 1980).

of IRT models and, as mentioned in the first part, expresses the probability of a correct response as a nonlinear function of two parameters simultaneously—a person's latent ability on the continuum of interest and the difficulty of the item, as expressed on that same continuum. Specifically, item difficulty is the location along a logistic function at which latent ability, as expressed along the *x*-axis, intersects with a 50% probability of a correct response, as expressed along the *y*-axis. All other possible item parameters (e.g., discrimination, or slope) are fixed. Latent ability is expressed as a single continuum, thus the need for demonstrating a unidimensional structure across all test items prior to IRT modeling (as done in Fergadiotis et al., 2023). Readers are directed to Embretson and Reise (2013) for a detailed introduction to the 1-PL and other IRT models.

For the present study, all 1-PL models were specified as generalized linear mixed-effects models (GLMMs), a general framework for modeling item response patterns (De Boeck & Wilson, 2004). These are all multivariable models with a single dependent variable and several covariates, expressed as either fixed or random effects. Specifically, the probability of a correct response, via a logit link function, is expressed as a linear function of items, persons, and their covariates (see Figure 1). In the absence of covariates, the equation is equivalent to a traditional 1-PL IRT model, as depicted in blue font in Figure 1. When covariates are included, as depicted in black font in Figure 1, they can be interpreted as effects on various components of the 1-PL IRT model, as described in greater detail below. This investigation of relations among person and/or item covariates in response to items is broadly referred to as *explanatory IRT*, whereas more traditional IRT models without covariates (i.e., measurement models) may be referred to as *descriptive IRT* (De Boeck & Wilson, 2004).

Figure 1. IRT modeling framework. Graphical depiction of the linear component of the final descriptive IRT model (1-PL IRT model with random effect for items), shown in blue font, and the addition of covariates for the final explanatory IRT model (latent regression LLTM), shown in black font, as expressed within a GLMM framework (notation adapted from the work of De Boeck & Wilson, 2004). The random and logit link function components that connect the observed VNT item responses to the logit-transformed probability of a correct response are not shown. GLMM = generalized linear mixed-effects model; IRT = item response theory; LLTM = linear logistic test model; VNT = Verb Naming Test; 1-PL = one-parameter logistic.



For the purposes of the present study, our first step was to identify an appropriate 1-PL IRT model without covariates (i.e., descriptive IRT model) that acted as a baseline model for subsequent modeling. Our second step was to expand on this baseline model to include item and person covariates and their interactions (i.e., explanatory IRT model). Following published guidelines (e.g., Cho et al., 2017), our final models were determined using likelihood ratio tests (LRTs) for nested model comparisons and two information criteria indices—Akaike information criterion (AIC; Akaike, 1973) and Bayesian information criterion (BIC; Schwarz, 1978). Proportion of variance explained (R^2) was also evaluated qualitatively. An overview of the research questions and related method of our two-step modeling approach is provided on the left-hand side of Figure 2.

Descriptive IRT Model (Step 1)

Building on our prior findings (Fergadiotis et al., 2023) and prior to investigating the three research questions of the present study, we tested whether the 1-PL IRT model adequately explained additional complexities within the data. Specifically, our first step (Step 1A) was to evaluate whether there was evidence of multilevel dependencies among the 18 AphasiaBank data collection sites (i.e., clusters) of the current data set. This was done by calculating the intraclass correlation coefficient, or the proportion of variance explained by cluster relative to the total variance, for nested 1-PL and multilevel 1-PL IRT models. Then (Step 1B), the best-fitting model was respecified with a random effect for items (as opposed to the traditional fixed effect), which allowed us to evaluate whether VNT items could be more productively understood as being sampled from a larger population of verb production items (De Boeck, 2008). This was done by inspecting the random effect's variance to determine if there was sufficient variability to warrant use of a random effect for items. This final step was completed to allow us to specify our explanatory IRT model, as described below, which necessitates a random item effect (i.e., unexplained variability over items) for unbiased estimates and standard errors of covariate effects (Cho et al., 2014; Janssen et al., 2004).

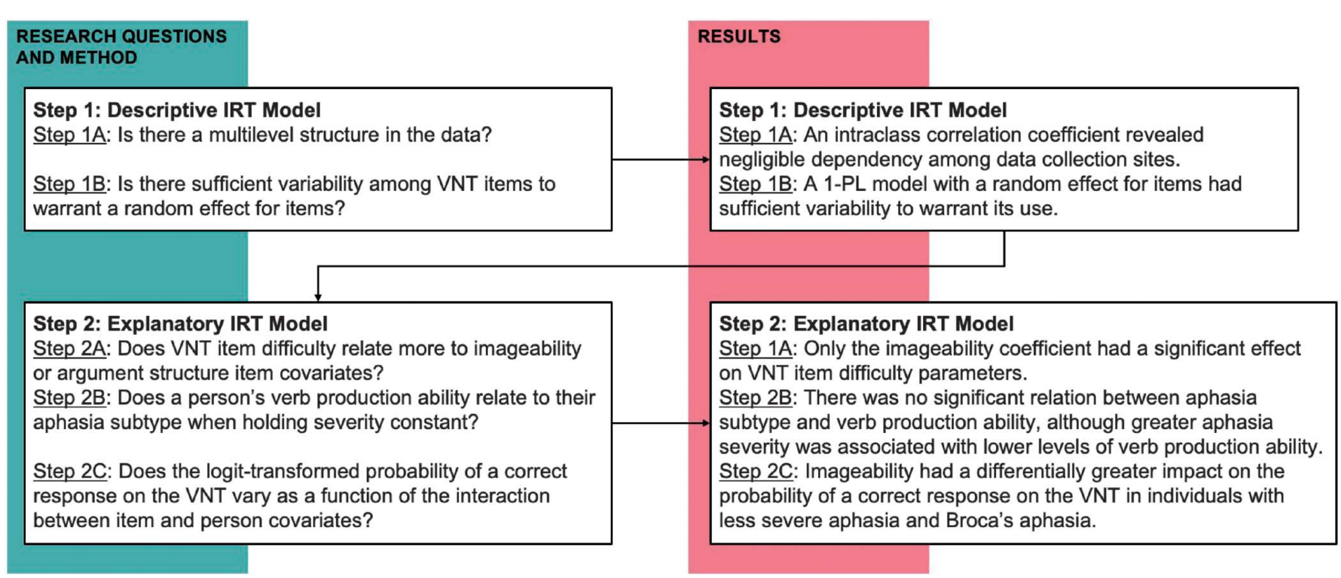
Explanatory IRT Model (Step 2)

To address our three research questions, we expanded upon our best-fitting descriptive IRT model to specify a single explanatory IRT model (latent regression linear logistic test model [LLTM]; see De Boeck & Wilson, 2004, for a detailed explanation of possible explanatory IRT models) that yields information about the effect of item covariates, person covariates, and Item × Person interactions on item response patterns of the VNT. All covariates were entered into the model as fixed; all other aspects of the final descriptive IRT model from Step 1 remained the same (see Figure 1). Details about how different effects within the model relate to the research questions at hand are outlined below.

Research Question 1: Explaining VNT Item Difficulty (Step 2A)

The inclusion of item covariates within a 1-PL IRT model permits the quantification of their effect on item

Figure 2. Summary of steps of the IRT modeling framework. Overview of the research questions and the method used to address them (left-hand side), as well as a brief outlining of the results (right-hand side). IRT = item response theory; VNT = Verb Naming Test; 1-PL = one-parameter logistic.



difficulty parameters. This "item side" in explanatory IRT is referred to as an LLTM with a random effect for items to account for unexplained variability (De Boeck, 2008) or as an item generation model (Cho et al., 2014). In the case of the present study, this LLTM model was first used for the preliminary analysis with three item covariates common to word production tests in aphasia, as mentioned above (see Supplemental Material S4 for details). Then, for our main analysis, we used the main item covariates (argument structure, imageability) to investigate our first research question, which pertained to identifying which cognitive process of word production was predominantly driving variability in VNT item difficulty.

Research Question 2: Explaining Verb Production Ability (Step 2B)

The "person side" of explanatory IRT, where the effect of person covariates on IRT test-level estimates of underlying latent ability (verb production), is commonly referred to as latent regression with a random effect for items (De Boeck & Wilson, 2004). Here, these covariates targeted our second research question, which was to identify whether individuals with the WAB-R subtype (Broca's) most associated with a morphosyntactic processing deficit (agrammatism) possessed differentially less verb production ability than those classified with Anomic or other subtypes when controlling for overall aphasia severity, as measured using AQ.

Research Question 3: Explaining VNT Response Probabilities (Step 2C)

The "person" and "item" aspects of explanatory IRT discussed thus far only covered main effects. However, our third research question pertained to the interaction of Item × Person covariates. Thus, we extended our explanatory IRT model (latent regression LLTM) to include interaction terms for testing whether the effect of an item covariate moderated the effect of a given person covariate on the probability of a correct response. Here, these interactions provide information about whether the effect of imageability or argument structure is contingent on a person's WAB-R subtype when controlling for their aphasia severity. This particular aspect of the model can also be referred to as differential facet functioning (Meulders & Xie, 2004) and provides information about the effect of Item × Person interactions on the probability of a correct response.

Results

All models converged to an admissible solution. A summary of results relative to each research question and its modeling approach is shown on the right-hand side of Figure 2.

Descriptive IRT Model (Step 1)

For investigating the multilevel structure of the data (Step 1A), we found that the intraclass correlation coefficient comparing variance from the random person effect between the 1-PL and multilevel 1-PL IRT models was .037, suggesting there was negligible variance among persons that was attributable to clusters (i.e., data collection site). As such, a 1-PL IRT model was retained as the best-fitting and more parsimonious model. Then, findings regarding the specification of a random effect for items (Step 1B) revealed an item variance of 0.978, which was judged to be sufficient to proceed with a random effect for items for the subsequent models based on prior work using similar types of IRT models (e.g., De Boeck et al., 2016). In summary, a 1-PL IRT model with random effect for items was identified as an appropriate baseline or descriptive IRT model to be expanded upon for the explanatory IRT model specification. Full model results for this final descriptive IRT model can be found in Supplemental Material S3.

Explanatory IRT Model (Step 2)

Relative to our final descriptive IRT model (1-PL IRT with random effect for items), the explanatory IRT model (latent regression LLTM), which included all of the person and item covariates and their interactions, demonstrated significantly better fit to the data. AIC and BIC values were lower, the LRT was significant ($p < .001$), and the model overall explained a substantial amount of both the item- and person-related variance (see Table 2). Findings specific to each research question are detailed below, with full results displayed in Table 3 and Figure 3.

Research Question 1: Explaining VNT Item Difficulty (Step 2A)

For the preliminary analysis including the traditional item covariates (lexical frequency, age of acquisition, and phoneme length), the LLTM with the covariates was not significantly different than the one without covariates, yielding a nonsignificant LRT ($p = .547$). In other words, the three item covariates combined did not explain any additional variability in item difficulty on the VNT (see Supplemental Material S4 for additional details).

In our main analysis, there was a significant main effect of imageability ($p < .001$) but not argument structure ($p = .195$) on VNT item difficulty parameters (see Table 3). For every 1 *SD* (0.80) increase in imageability, item difficulty parameters decreased on average by 0.935

Table 2. Model fit indices for the descriptive and explanatory IRT models.

Model	Number of parameters	AIC	BIC	Log-likelihood	Deviance	LRT			Item R^2	Person R^2
						χ^2	*df*	*p*		
1-PL IRT with random item effect	3	2608.7	2626.0	−1301.3	2602.7	NA	NA	NA	NA	NA
Latent regression LLTM	14	2466.7	2547.3	−1219.3	2438.7	164.04	11	< .001	.706	.772

Note. IRT = item response theory; AIC = Akaike information criterion; BIC = Bayesian information criterion; LRT = likelihood ratio test; 1-PL = one-parameter logistic; LLTM = linear logistic test model.

on the logit scale, equivalent to an odds ratio decrease of 0.393, while holding other covariates constant. This corresponds to a ~23% decrease on the probability scale when comparing item difficulty parameters that lie between approximately −1 and +1 on the logit scale (implying a range of approximately 25%–75% correct on the probability scale). In other words, the difficulty of an item on the VNT increased as the targeted verb became less imageable.

Research Question 2: Explaining Verb Production Ability (Step 2B)

There was a significant main effect for the controlling covariate of AQ ($p < .001$) but not for subtype levels, either Anomic versus Broca's ($p = .262$) or Other versus Broca's ($p = .464$) on verb production ability (see Table 3). For every 1 *SD* increase in AQ (17.1), verb production ability increased by 1.387 on the logit scale, equivalent to an odds ratio increase of 4.003, while holding other covariates constant. This corresponds to a ~34% increase on the probability scale when comparing individuals whose verb production abilities lies between approximately −1 and +1 on the logit scale. In other words, individuals with higher AQ or less severe aphasia severity had higher verb production ability on average.

Research Question 3: Explaining VNT Response Probabilities (Step 2C)

In addition to the main effects discussed above and as shown in Table 3, there were two significant interaction effects: Imageability × AQ ($p = .03$) and Imageability × Subtype, Anomic versus Broca's ($p = .021$). Plots of the simple slopes revealed that imageability had a differentially

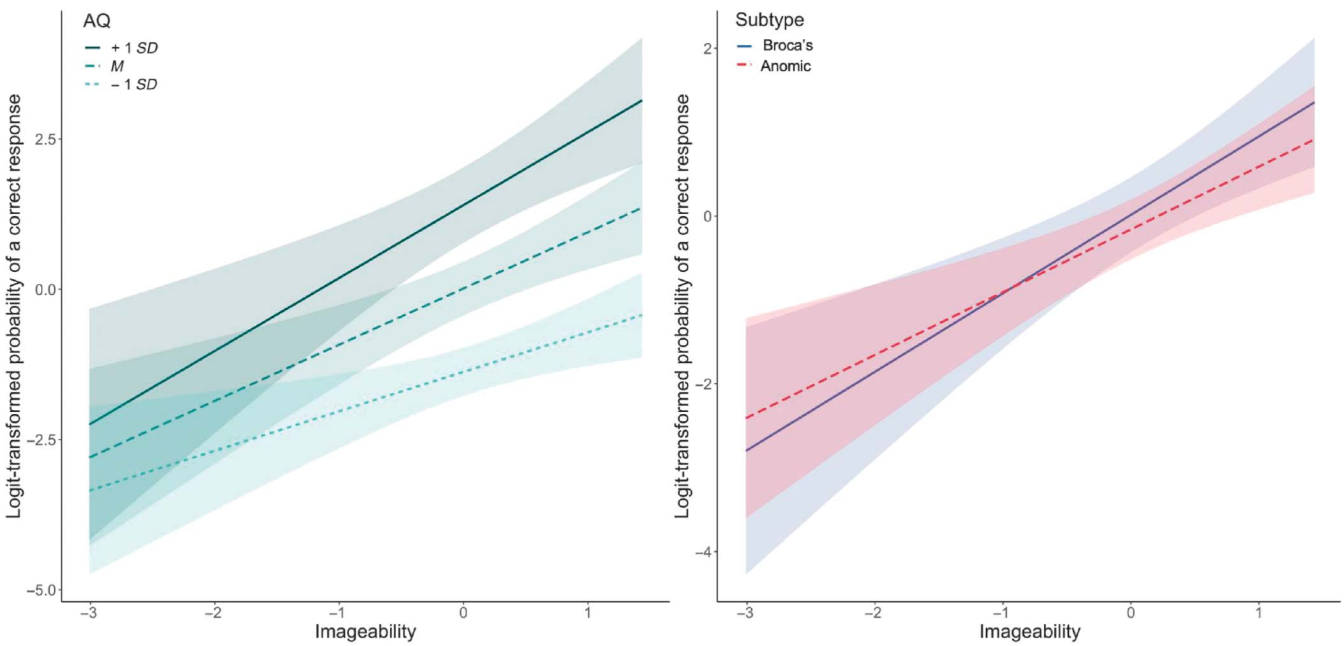
Table 3. Fixed and random effects of the explanatory IRT model (latent regression LLTM).

Fixed effects	Estimate (logit)	Estimate (odds ratio)	SE^a	*z*-statistic	*p*-value
Intercept	0.015	1.015	0.229	0.063	.950
Argument structure	0.274	1.315	0.211	−1.297	.195
Imageability	−0.935	0.393	0.233	4.012	< .001
Subtype, anomic versus Broca's	−0.358	0.699	0.319	−1.122	.262
Subtype, other versus Broca's	−0.173	0.841	0.237	−0.732	.464
AQ	1.387	4.003	0.141	9.828	< .001
Argument structure*Subtype, anomic versus Broca's	−0.110	0.896	0.249	−0.442	.659
Argument structure*Subtype, other versus Broca's	0.153	1.165	0.183	0.837	.403
Argument structure*AQ	−0.053	0.948	0.199	−0.446	.656
Imageability*Subtype, anomic versus Broca's	−0.642	0.526	0.279	−2.303	.021
Imageability*Subtype, other versus Broca's	−0.187	0.829	0.213	−0.879	.379
Imageability*AQ	0.278	1.320	0.128	2.171	.030
Random effects	Variance				
Person	0.489				
Item	0.287				

Note. Transformation of parameter estimates from the logit scale to an odds ratio was calculated using the following formula: exp(logit). Item covariate effect estimates, when specified using the "glmer" function in the *lme4* package in R (Bates et al., 2015), are expressed on an "easiness" scale, as opposed to the traditional "difficulty" scale of IRT models; as such, value signs were flipped to re-express estimates on the "difficulty" scale for ease of interpretation. IRT = item response theory; LLTM = linear logistic test model; AQ = Aphasia Quotient.

[a]*SE* = logit-scaled standard error.

Figure 3. Significant interactions from the latent regression LLTM. Plots depicting the significant interaction between imageability and AQ (left) and between imageability and subtype, anomic versus Broca's (right). Solid and dashed lines show the logit-transformed predicted VNT item responses as a function of the imageability item covariate for each quantitative (AQ) and qualitative (subtype) level of the person covariates when holding all other covariates constant. The shaded bands represent the 95% confidence interval around the predicted slopes. AQ = Aphasia Quotient; LLTM = linear logistic test model; VNT = Verb Naming Test.



greater impact on the logit-transformed probability of a correct response on the VNT in individuals with higher AQs, or less severe aphasia, and those classified with the Broca's subtype, as compared with the Anomic subtype when holding all other covariates constant (see Figure 3).

Discussion

The aim of the present study was to evaluate the underlying constructs of the VNT by expanding and refining our previously reported IRT modeling framework (Fergadiotis et al., 2023). In a diverse sample of 107 participants with aphasia, we found that the item response patterns on the VNT were primarily associated with item (imageability) and person (WAB-R AQ) covariates, as well as their interaction, thought to reflect lexical-semantic processing. In contrast, argument structure and the Broca's WAB-R subtype—respective item and person covariates that have previously been assumed to reflect morphosyntactic processing (Cho-Reyes & Thompson, 2012)—were not explanatory of VNT item response patterns. However, imageability had a differentially greater impact on the probability of a correct response for those with the Broca's WAB-R subtype. Thus, when considering all findings collectively, the enacted construct of the VNT (i.e., lexical-semantic processing) does not clearly align with the test developers' intended construct (i.e., morphosyntactic processing; Cho-Reyes & Thompson, 2012; Gorin & Embretson, 2006).

The main findings of the present study build on a robust descriptive IRT model we evaluated in our prior work (Fergadiotis et al., 2023) and was refined herein. Notably, we confirmed the appropriateness of treating item difficulty as a random parameter and also identified no multilevel structure to the VNT item responses as a function of testing group site. This latter point suggests that standardized word production tests of aphasia similar to the VNT may be relatively invariant to clinician or geographic location, a consideration of particular importance given the widespread use of data from multisite repositories such as AphasiaBank (MacWhinney et al., 2011) for research purposes.

VNT Item Difficulty Is Modulated by Imageability (Research Question 1)

Our preliminary LLTM analysis revealed no combined effect of lexical frequency, age of acquisition, and phoneme length—item covariates reflecting both lexical-semantic and phonological processing—on VNT item difficulty parameters. This was to be expected given that these covariates are relatively controlled for in the VNT test design, in contrast with other confrontation naming tests of noun production, where these properties are purposefully manipulated (e.g., Fergadiotis, Swiderski, et al., 2019).

In the LLTM portion of our explanatory IRT model (latent regression LLTM), imageability, a covariate representative of lexical-semantic processing, was highly associated with item difficulty parameters even though it was not considered in the VNT design process. Although imageability has received relatively less attention in the aphasia literature as compared with other item covariates (Lampe et al., 2021; Nickels, 1995; Nickels et al., 2022), its important role in the present study comports not only with research on word production performance in healthy individuals (Perret & Bonin, 2019) but also those with aphasia, where no group-level dissociations between noun and verb production performance are observed when imageability ratings, among other item covariates (e.g., lexical frequency), are matched across stimuli or otherwise accounted for in study analyses (Aggujaro et al., 2006; Alyahya et al., 2018; Berndt et al., 2002; Bird et al., 2000; Bird, Howard, et al., 2003; Crepaldi et al., 2006; Luzzatti et al., 2002; Mätzig et al., 2009; Vigliocco et al., 2011). Notably, some of these group studies identified a small number of participants who persisted in having a differentially greater impairment with verbs even when imageability was controlled for (e.g., Aggujaro et al., 2006; Crepaldi et al., 2006; Mätzig et al., 2009), as has been observed in case reports in the neuropsychological literature (e.g., Miceli et al., 1984). Thus, our finding on imageability does not negate the possibility of true grammatical class dissociations among the participants included in this study or the aphasia population at large but rather suggests that such dissociations are uncommon and require careful behavioral testing.

Argument structure, a covariate that is assumed to be representative of verb-related morphosyntactic processing, was the primary item property manipulated in the VNT, yet we identified no significant effect of argument structure on item difficulty parameters. The underlying reasons for this are likely multifactorial. First, argument structure is inherently more restricted in its range (1–3) than imageability (2.47–5.92) among the 22 VNT items, thus potentially yielding less information. Second, argument structure can be challenging to quantify, and it is not uncommon for raters to disagree on the number of the participant roles for the same verb (FitzGerald et al., 2018). Third, there is some degree of visual anomaly among the picture stimuli of the VNT, where a subset of verbs not explicitly obligatory in their argument structure was forced to be as such through the inclusion of additional elements (Cho-Reyes & Thompson, 2012). Although visual complexity as an item covariate is both distinct from imageability and seemingly unimportant to word production (Nickels et al., 2022; Perret & Bonin, 2019), idiosyncratic alternations such as these may nonetheless have had unintended consequences and introduced construct-irrelevant noise.

Finally, argument structure is itself a complex property spanning both morphosyntactic and lexical-semantic processes, and the contributions from each can be challenging to distinguish. Regarding morphosyntactic processing, argument structure plays a critical role in not only specifying the number and type of object phrases but also the grammatical morphology needed for each object phrase (Grimshaw, 1990; Shapiro et al., 1987). Regarding lexical-semantic processing, argument structure carries information about underlying semantic relationships (Dowty, 1991; Levin, 1993; Pinker, 2013); can be primed by semantic information (Ferretti et al., 2001; Hare et al., 2009; McRae et al., 2005); and is systematically related to properties of lexical-semantic processing, such as semantic weight (i.e., the number of distinctive semantic features a verb has; Gordon & Dell, 2003). Argument structure differences between verbs are therefore just as likely to reflect lexical-semantic as morphosyntactic processes and representations. Relevant to this point, there was a moderate correlation (r = −.625, 95% confidence interval [−.348, −.780][3]) between argument structure and imageability, which itself is highly correlated with semantic weight (Barde et al., 2006), among the VNT items. Thus, when considered in conjunction with the other limitations noted earlier, it is likely that any potential unique effect of argument structure, if present, is not distinguishable from that of imageability or other lexical-semantic features. This broader challenge of isolating morphosyntactic processing from that of lexical-semantics is one that is well established and, recently, has received considerable attention in the neuroimaging literature (e.g., Fedorenko et al., 2020; Hu et al., 2022; Matchin & Hickok, 2020).

Verb Production Ability Is Modulated by Aphasia Severity (Research Question 2)

For the latent regression portion of our explanatory IRT model, our controlling covariate of aphasia severity, as measured with the WAB-R AQ, was a significant predictor of verb production ability on the VNT. This finding aligns with a larger body of literature showing the close alignment between overall word production ability and aphasia severity (e.g., Kristinsson et al., 2023; Walker & Schwartz, 2012; Wilson et al., 2023), as mentioned in the

[3]This confidence interval was calculated via a bias-corrected accelerated bootstrap method (DiCiccio & Efron, 1996; Efron, 1987) from 10,000 iterations of resampled data using the *boot* package (Canty & Ripley, 2017) in R (Version 4.3.0).

first part. It also comports with research showing that impaired verb production ability, like impaired noun production ability, is not only highly prevalent among people with aphasia (Mätzig et al., 2009) but also associated with overall functional communication in aphasia (Rofes et al., 2015).

Aphasia subtype, again as measured using the WAB-R, had no significant effect on verb production ability. Most importantly, there was no difference in verb production ability between the Broca's (agrammatic) and Anomic subtypes. This contrasts with dissociations reported between these two subgroups in the initial psychometric validation of the VNT (Cho-Reyes & Thompson, 2012), among other studies (e.g., Lee & Thompson, 2015; Zingeser & Berndt, 1990). In line with our earlier finding on item covariates, we interpret this lack of effect as originating from multiple sources. First, the absence of aphasia subtype differences is likely derivative, in part, of the VNT's item design. In other words, given that an item covariate assumed to be relevant to morphosyntactic processing (i.e., argument structure) was not associated with item-specific parameters (i.e., difficulty), it is unlikely that a person covariate assumed to be relevant to morphosyntactic processing (i.e., Broca's or agrammatic subtype) would be associated with person-specific parameters (i.e., verb production ability). Second, our study's sample was much larger and more varied than previous work showing subtype-specific effects, containing 107 participants with no restriction on the basis of aphasia subtype. Other studies with samples similar to ours (e.g., Alyahya et al., 2018) have also failed to identify grammatical class dissociations or meaningful groupings on the basis of aphasia subtypes, thus suggesting that such dissociations are either rare or potentially an artifact of sampling bias. Third and finally, Broca's aphasia is a relatively amorphous syndrome that is only putatively associated with agrammatism, as outlined in the first part. Moreover, all people with aphasia are anomic to varying degrees (Goodglass & Wingfield, 1997), yet the subtype as defined within the WAB-R is far more restrictive (Kertesz, 2006). We have emphasized these subtypes in the present study here given their historical precedent in the literature and on the VNT and NAVS specifically. However, these findings further illustrate the problems associated with subtypes, which appear to obscure more than elucidate our understanding of the underlying nature of aphasia (Alyahya et al., 2018; E. Bates et al., 2005; Butler et al., 2014; Casilio et al., 2019; Landrigan et al., 2021; Wilson et al., 2023).

VNT Response Probabilities Are Differentially Modulated by Imageability for Those With Less Severe or Broca's Aphasia (Research Question 3)

For the latent regression LLTM portion of our explanatory IRT model, imageability was a significant moderator of both aphasia severity (WAB-R AQ) and subtype (WAB-R subtype). For the interaction between imageability and severity, individuals with less severe impairments were more sensitive to imageability in their probability of producing a correct response. This finding is to be expected given that those with more severe aphasia likely have impairments across multiple cognitive processes important to word production (e.g., Schwartz & Brecher, 2000), thus making their response probabilities on the VNT relatively invariant to more subtle changes in a single item covariate.

The interaction between imageability and subtype may be consistent with previous findings showing that verb production in those with agrammatism is robustly affected by properties indexing lexical-semantic processing (Barde et al., 2006; Park et al., 2023; Rezaii et al., 2023; Thorne & Faroqi-Shah, 2016). Specifically, semantically heavy verbs (e.g., "fly") are produced at equal or in greater number, whereas semantically light verbs (e.g., "go") are produced in smaller number during elicited sentence production and connected speech tasks. This association is analogous to the one identified here—individuals with Broca's subtype were more likely to accurately produce more imageable verbs. This interpretation should be taken with caution, given the imperfect relationship between the Broca's subtype and agrammatism, as outlined above. Nonetheless, neurobiological evidence shows that those classified with both the Broca's subtype (Mohr, 1976; Mohr et al., 1978) and agrammatism (Matchin et al., 2020) commonly have lesions sparing the left temporal lobe and consequently have preserved access to lexical and (critically) semantic representations stored there (e.g., Hickok & Poeppel, 2007; Lambon Ralph et al., 2017; Wilson et al., 2009, 2018). Thus, this preserved access may make response probabilities on the VNT more sensitive to manipulations in imageability, potentially as a form of strategic compensation (Fedorenko et al., 2022; Gordon & Dell, 2003). Such an interpretation would comport with findings from the behavioral literature showing people with aphasia, as compared with healthy speakers, make greater use of semantic cues during verb production (Dresang et al., 2021).

In contrast, there was no moderating effect of argument structure on aphasia severity or subtype. Although the prior literature has shown that those with agrammatic aphasia tend to be more sensitive to alterations in argument structure (Cho-Reyes & Thompson, 2012; Kim & Thompson, 2000; Thompson, 2003), this lack of effect is congruent with other findings from the explanatory IRT model and likely a byproduct of the overall test design, among other factors, as discussed above.

Clinical and Research Applications

There are multiple potential ways in which the findings of the current study could apply to future clinical and research endeavors. Specifically, the parameter estimates (item difficulty, verb production ability) of the final descriptive IRT model, as reported in Supplemental Material S3, could be used to evaluate relations among VNT performance and other available AphasiaBank data using psychometrically robust test- and item-level covariates. Moreover, the item difficulty parameter estimates could be used to derive estimates of verb production ability in new groups of participants assessed with the VNT. Additionally, with regard to the explanatory IRT model, the set of covariates we evaluated could be leveraged to develop a predictive model for estimating item difficulty of new items for the purposes of refining or expanding on the VNT. Finally, the current findings further emphasize the importance of item covariates like imageability, among others (e.g., lexical frequency), in assessing verb production ability in aphasia and, more broadly, in developing a diagnostic profile of performance to guide treatment planning. Given that the cognitive constructs underlying many aphasia tests are presumed as opposed to empirically tested, we urge caution for both clinicians and researchers when using not only VNT but other similar tests that purport to test verb-related morphosyntactic processing in aphasia. Specifically, we recommend the VNT be used primarily as a measure of lexical-semantic processing. Although the VNT may capture some degree of information about morphosyntactic processing, this should be corroborated with other measures known to capture behaviors salient to this process, such as those derived from a connected speech sample (see Saffran et al., 1989, and Casilio et al., 2019, as examples of validated transcription-based and transcription-less connected speech analysis systems). In our view, other confrontation naming tests of verb production that do not control for imageability are likely to also predominantly capture lexical-semantic processing, in line with other empirical studies of verb production (e.g., Alyahya et al., 2018). Whether our findings pertaining to the VNT extend to the larger NAVS, however, remains an open question, and future research should explore the constructs underlying other subtests in the battery.

Limitations and Future Directions

The present study is not without limitations. First, motor planning and execution are processes known to be important to successful word production (e.g., Walker & Hickok, 2016). Although the VNT scoring criteria control for these processes to the extent possible (Cho-Reyes & Thompson, 2012), we were unable to evaluate whether relevant person covariates (i.e., clinical diagnosis of apraxia of speech or dysarthria) were associated with VNT item response patterns due to insufficient power. Future research on verb production is ongoing among our research team, and the role of motoric processes in word production using IRT is planned in future once sample sizes permit. Second, our findings focused on a single verb production test for aphasia, and future research should ideally incorporate multiple tests of both verbs and nouns to better elucidate the cognitive processes underlying each and in relation to one another. Third and finally, measures of structural brain damage are known to be explanatory of verb production in aphasia (e.g., Aggujaro et al., 2006; Alyahya et al., 2018; Dresang et al., 2021). Incorporating them into a modeling framework would likely yield additional information regarding the VNT's underlying constructs and thus would be a valuable future direction.

Conclusions

Although verb production may lie at the intersection of lexical-semantic and morphosyntactic processes, many verb production tests are insufficiently validated for testing this claim. Our careful analysis of item response patterns using IRT, as obtained from a diverse group of people with aphasia, suggests that the underlying construct of the VNT, a commonly used verb production test in aphasia, is most associated with covariates relevant to lexical-semantic processing. These findings illustrate the dual nature of argument structure as both a morphosyntactic and lexical-semantic verb property; reiterate the importance of controlling for imageability, among other relevant item covariates, in studies pertaining to the cognitive processes underlying word production in aphasia; and recapitulate the particular challenge of isolating morphosyntactic processing from lexical and semantic processes.

Data Availability Statement

Audiovisual recordings and participant-level demographic and behavioral testing information are freely available on AphasiaBank (https://aphasia.talkbank.org). Scored item responses on the VNT, along with the covariates extracted for the analyses of the present study, are included in Supplemental Material S1. Analysis code is available in Supplemental Material S2.

Acknowledgments

The authors gratefully acknowledge the support of National Institute on Deafness and Other Communication

Disorders Awards R01DC018813 (principal investigators [PIs]: William Hula and Gerasimos Fergadiotis) and F31DC021108 (PI: Marianne Casilio), as well as a PhD scholarship from the Council on Academic Programs in Communication Sciences and Disorders (PI: Marianne Casilio). The authors would also like to thank the AphasiaBank participants who generously donated their time, as well as the research assistants at Portland State University for their support in the transcription and scoring of the data set used in the present study.

References

Aggujaro, S., Crepaldi, D., Pistarini, C., Taricco, M., & Luzzatti, C. (2006). Neuro-anatomical correlates of impaired retrieval of verbs and nouns: Interaction of grammatical class, imageability and actionality. *Journal of Neurolinguistics, 19*(3), 175–194. https://doi.org/10.1016/j.jneuroling.2005.07.004

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika, 60*(2), 255–265. https://doi.org/10.1093/biomet/60.2.255

Alyahya, R. S. W., Halai, A. D., Conroy, P., & Lambon Ralph, M. A. (2018). Noun and verb processing in aphasia: Behavioural profiles and neural correlates. *NeuroImage: Clinical, 18,* 215–230. https://doi.org/10.1016/j.nicl.2018.01.023

Barde, L. H., Schwartz, M. F., & Boronat, C. B. (2006). Semantic weight and verb retrieval in aphasia. *Brain and Language, 97*(3), 266–278. https://doi.org/10.1016/j.bandl.2005.11.002

Basso, A., Razzano, C., Faglioni, P., & Zanobio, M. E. (1990). Confrontation naming, picture description and action naming in aphasic patients. *Aphasiology, 4*(2), 185–195. https://doi.org/10.1080/02687039008249069

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), https://doi.org/10.18637/jss.v067.i01

Bates, E., Chen, S., Tzeng, O., Li, P., & Opie, M. (1991). The noun–verb problem in Chinese aphasia. *Brain and Language, 41*(2), 203–233. https://doi.org/10.1016/0093-934X(91)90153-R

Bates, E., Saygin, A. P., Moineau, S., Marangolo, P., & Pizzamiglio, L. (2005). Analyzing aphasia data in a multidimensional symptom space. *Brain and Language, 92*(2), 106–116. https://doi.org/10.1016/j.bandl.2004.06.108

Benson, D. F. (1967). Fluency in aphasia: Correlation with radioactive scan localization. *Cortex, 3*(4), 373–394. https://doi.org/10.1016/S0010-9452(67)80025-X

Benson, D. F. (1979). *Aphasia, alexia, and agraphia*. Churchill Livingstone.

Berndt, R. S., Haendiges, A. N., Burton, M. W., & Mitchum, C. C. (2002). Grammatical class and imageability in aphasic word production: Their effects are independent. *Journal of Neurolinguistics, 15*(3-5), 353–371. https://doi.org/10.1016/S0911-6044(01)00030-6

Bird, H., Lambon Ralph, M. A., Seidenberg, M. S., McClelland, J. L., & Patterson, K. (2003). Deficits in phonology and past-tense morphology: What's the connection? *Journal of Memory and Language, 48*(3), 502–526. https://doi.org/10.1016/S0749-596X(02)00538-7

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers, 33*(1), 73–79. https://doi.org/10.3758/BF03195349

Bird, H., Howard, D., & Franklin, S. (2003). Verbs and nouns: The importance of being imageable. *Journal of Neurolinguistics, 16*(2–3), 113–149. https://doi.org/10.1016/S0911-6044(02)00016-7

Bird, H., Lambon Ralph, M. A., Patterson, K., & Hodges, J. R. (2000). The rise and fall of frequency and imageability: Noun and verb production in semantic dementia. *Brain and Language, 73*(1), 17–49. https://doi.org/10.1006/brln.2000.2293

Black, M., & Chiat, S. (2003). Noun–verb dissociations: A multi-faceted phenomenon. *Journal of Neurolinguistics, 16*(2-3), 231–250. https://doi.org/10.1016/S0911-6044(02)00017-9

Butler, R. A., Lambon Ralph, M. A., & Woollams, A. M. (2014). Capturing multidimensionality in stroke aphasia: Mapping principal behavioural components to neural structures. *Brain, 137*(12), 3248–3266. https://doi.org/10.1093/brain/awu286

Canty, A., & Ripley, B. (2017). *boot: Bootstrap R (S-Plus) functions* (Version 1.3–20). R Foundation for Statistical Computing.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature, 349*(6312), 788–790. https://doi.org/10.1038/349788a0

Casilio, M., Kasdan, A. V., Bryan, K., Shibata, K., Schneck, S., Levy, D. F., Entrup, J., Onuscheck, C. F., de Riesthal, M., & Wilson, S. M. (2025). Four dimensions of naturalistic language in aphasia after stroke. *Brain, 148*(1), 291–312. https://doi.org/10.1093/brain/awae195

Casilio, M., Rising, K., Beeson, P. M., Bunton, K., & Wilson, S. M. (2019). Auditory-perceptual rating of connected speech in aphasia. *American Journal of Speech-Language Pathology, 28*(2), 550–568. https://doi.org/10.1044/2018_AJSLP-18-0192

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Cho, S.-J., De Boeck, P., Embretson, S., & Rabe-Hesketh, S. (2014). Additive multilevel item structure models with random residuals: Item modeling for explanation and item generation. *Psychometrika, 79*(1), 84–104. https://doi.org/10.1007/s11336-013-9360-2

Cho, S.-J., De Boeck, P., & Lee, W.-Y. (2017). Evaluating testing, profile likelihood confidence interval estimation, and model comparisons for item covariate effects in linear logistic test models. *Applied Psychological Measurement, 41*(5), 353–371. https://doi.org/10.1177/0146621617692078

Cho-Reyes, S., & Thompson, C. K. (2012). Verb and sentence production and comprehension in aphasia: Northwestern Assessment of Verbs and Sentences (NAVS). *Aphasiology, 26*(10), 1250–1277. https://doi.org/10.1080/02687038.2012.693584

Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 371–383. https://doi.org/10.3758/BF03195584

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods, Instruments, & Computers, 36*(3), 384–387. https://doi.org/10.3758/BF03195585

Crepaldi, D., Aggujaro, S., Arduino, L. S., Zonca, G., Ghirardi, G., Inzaghi, M. G., Colombo, M., Chierchia, G., & Luzzatti, C. (2006). Noun–verb dissociation in aphasia: The role of imageability and functional locus of the lesion. *Neuropsychologia, 44*(1), 73–89. https://doi.org/10.1016/j.neuropsychologia.2005.04.006

De Bleser, R., & Kauschke, C. (2003). Acquisition and loss of nouns and verbs: Parallel or divergent patterns? *Journal of*

Neurolinguistics, 16(2-3), 213–229. https://doi.org/10.1016/S0911-6044(02)00015-5
De Boeck, P. (2008). Random item IRT models. *Psychometrika, 73*(4), 533–559. https://doi.org/10.1007/s11336-008-9092-x
De Boeck, P., Cho, S.-J., & Wilson, M. (2016). Explanatory item response models: An approach to cognitive assessment. In A. Rupp & J. Leighton (Eds.), *Handbook of Cognition and Assessment* (pp. 249–266). Wiley Blackwell. https://doi.org/10.1002/9781118956588.ch11
De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. Springer. https://doi.org/10.1007/978-1-4757-3990-9
Dell, G. S., Lawler, E. N., Harris, H. D., & Gordon, J. K. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology, 21*(2–4), 125–145. https://doi.org/10.1080/02643290342000320
Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801
DiCiccio, T., & Efron, B. (1996). Bootstrap confidence intervals (with discussion). *Journal of the American Statistical Association, 11,* 189–228.
Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67*(3), 547–619. https://doi.org/10.1353/lan.1991.0021
Dresang, H. C., Hula, W. D., Yeh, F.-C., Warren, T., & Dickey, M. W. (2021). White-matter neuroanatomical predictors of aphasic verb retrieval. *Brain Connectivity, 11*(4), 319–330. https://doi.org/10.1089/brain.2020.0921
Dresang, H. C., Warren, T., Hula, W. D., & Dickey, M. W. (2021). Rational adaptation in using conceptual versus lexical information in adults with aphasia. *Frontiers in Psychology, 12,* Article 589930. https://doi.org/10.3389/fpsyg.2021.589930
Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82*(397), 171–185. https://doi.org/10.1080/01621459.1987.10478410
Embretson, S. E., & Reise, S. P. (2013). *Item response theory for psychologists*. Psychology Press. https://doi.org/10.4324/9781410605269
Faroqi-Shah, Y., & Dickey, M. W. (2009). On-line processing of tense and temporality in agrammatic aphasia. *Brain and Language, 108*(2), 97–111. https://doi.org/10.1016/j.bandl.2008.10.003
Faroqi-Shah, Y., & Thompson, C. K. (2007). Verb inflections in agrammatic aphasia: Encoding of tense features. *Journal of Memory and Language, 56*(1), 129–151. https://doi.org/10.1016/j.jml.2006.09.005
Fedorenko, E., Blank, I. A., Siegelman, M., & Mineroff, Z. (2020). Lack of selectivity for syntax relative to word meanings throughout the language network. *Cognition, 203,* Article 104348. https://doi.org/10.1016/j.cognition.2020.104348
Fedorenko, E., Ryskin, R., & Gibson, E. (2022). Agrammatic output in non-fluent, including Broca's, aphasia as a rational behavior. *Aphasiology, 37*(12), 1981–2000. https://doi.org/10.1080/02687038.2022.2143233
Fergadiotis, G., Casilio, M., Dickey, M. W., Steel, S., Nicholson, H., Fleegle, M., Swiderski, A., & Hula, W. D. (2023). Item response theory modeling of the Verb Naming Test. *Journal of Speech, Language, and Hearing Research, 66*(5), 1718–1739. https://doi.org/10.1044/2023_JSLHR-22-00458
Fergadiotis, G., Casilio, M., Hula, W. D., & Swiderski, A. (2021). Computer adaptive testing for the assessment of anomia severity. *Seminars in Speech and Language, 42*(03), 180–191. https://doi.org/10.1055/s-0041-1727252
Fergadiotis, G., Hula, W. D., Swiderski, A. M., Lei, C.-M., & Kellough, S. (2019). Enhancing the efficiency of confrontation naming assessment for aphasia using computer adaptive testing. *Journal of Speech, Language, and Hearing Research, 62*(6), 1724–1738. https://doi.org/10.1044/2018_JSLHR-L-18-0344
Fergadiotis, G., Kellough, S., & Hula, W. D. (2015). Item response theory modeling of the Philadelphia Naming Test. *Journal of Speech, Language, and Hearing Research, 58*(3), 865–877. https://doi.org/10.1044/2015_JSLHR-L-14-0249
Fergadiotis, G., Swiderski, A., & Hula, W. D. (2019). Predicting confrontation naming item difficulty. *Aphasiology, 33*(6), 689–709. https://doi.org/10.1080/02687038.2018.1495310
Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language, 44*(4), 516–547. https://doi.org/10.1006/jmla.2000.2728
FitzGerald, N., Michael, J., He, L., & Zettlemoyer, L. (2018). *Large-scale QA-SRL parsing*. arXiv. https://doi.org/10.48550/arXiv.1805.05377
Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43*(2), 182–216. https://doi.org/10.1006/jmla.2000.2716
Franklin, S., Howard, D., & Patterson, K. (1995). Abstract word anomia. *Cognitive Neuropsychology, 12*(5), 549–566. https://doi.org/10.1080/02643299508252007
Friedmann, N., & Grodzinsky, Y. (1997). Tense and agreement in agrammatic production: Pruning the syntactic tree. *Brain and Language, 56*(3), 397–425. https://doi.org/10.1006/brln.1997.1795
Gao, C., Shinkareva, S. V., & Desai, R. H. (2022). SCOPE: The South Carolina psycholinguistic metabase. *Behavior Research Methods, 55*(6), 2853–2884. https://doi.org/10.3758/s13428-022-01934-0
Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation, 12*(4), 395–427. https://doi.org/10.3758/BF03201693
Goodglass, H. (1997). Agrammatism in aphasiology. *Clinical Neuroscience, 4*(2), 51–56.
Goodglass, H., & Wingfield, A. (1997). *Anomia: Neuroanatomical and cognitive correlates*. The Academic Press.
Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science, 27*(1), 1–40. https://doi.org/10.1207/s15516709cog2701_1
Gorin, J. S., & Embretson, S. E. (2006). Item difficulty modeling of paragraph comprehension items. *Applied Psychological Measurement, 30*(5), 394–411. https://doi.org/10.1177/0146621606288554
Graves, W. W., Desai, R., Humphries, C., Seidenberg, M. S., & Binder, J. R. (2010). Neural systems for reading aloud: A multiparametric approach. *Cerebral Cortex, 20*(8), 1799–1815. https://doi.org/10.1093/cercor/bhp245
Grimshaw, J. (1990). *Argument structure*. MIT Press.
Grimshaw, J., & Jackendoff, R. (1981). *Brandeis verb lexicon*. Electronic Database Funded by National Science Foundation Grant NSF IST-81-20403 Awarded to Brandeis University.
Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition, 111*(2), 151–167. https://doi.org/10.1016/j.cognition.2009.01.009
Henning, G. (1992). Dimensionality and construct validity of language tests. *Language Testing, 9*(1), 1–11. https://doi.org/10.1177/026553229200900102

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402. https://doi.org/10.1038/nrn2113
Horton, M., Marais, I., & Christensen, K. B. (2012). Dimensionality. In K. B. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch models in health* (pp. 137–158). Wiley. https://doi.org/10.1002/9781118574454.ch9
Hu, J., Small, H., Kean, H., Takahashi, A., Zekelman, L., Kleinman, D., Ryan, E., Nieto Castañón, A., Ferreira, V., & Fedorenko, E. (2022). Precision fMRI reveals that the language-selective network supports both phrase-structure building and lexical access during language production. *Cerebral Cortex, 33*(8), 4384–4404. https://doi.org/10.1093/cercor/bhac350
Hula, W. D., Fergadiotis, G., Swiderski, A. M., Silkes, J. P., & Kellough, S. (2020). Empirical evaluation of computer-adaptive alternate short forms for the assessment of anomia severity. *Journal of Speech, Language, and Hearing Research, 63*(1), 163–172. https://doi.org/10.1044/2019_JSLHR-L-19-0213
Hula, W. D., Kellough, S., & Fergadiotis, G. (2015). Development and simulation testing of a computerized adaptive version of the Philadelphia Naming Test. *Journal of Speech, Language, and Hearing Research, 58*(3), 878–890. https://doi.org/10.1044/2015_JSLHR-L-14-0297
Janssen, R., Schepers, J., & Peres, D. (2004). Models with item and item group predictors. In P. De Boeck, & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 189–212). Springer. https://doi.org/10.1007/978-1-4757-3990-9_6
Johnson, C. J., Paivio, A., & Clark, J. M. (1996). Cognitive components of picture naming. *Psychological Bulletin, 120*(1), 113–139. https://doi.org/10.1037/0033-2909.120.1.113
Kaplan, E., Goodglass, H., & Weintraub, S. (2001). *Boston Naming Test* (2nd ed.). Lippincott Williams & Wilkins.
Kertesz, A. (1979). *Aphasia and associated disorders: Taxonomy, localization, and recovery*. Holt Rinehart & Winston.
Kertesz, A. (2006). *Western Aphasia Battery—Revised*. Grune and Stratton.
Kim, M., & Thompson, C. K. (2000). Patterns of comprehension and production of nouns and verbs in agrammatism: Implications for lexical organization. *Brain and Language, 74*(1), 1–25. https://doi.org/10.1006/brln.2000.2315
Kiss, K. (1997). The effect of verb complexity on agrammatic aphasics' sentence production. *Acta Linguistica Hungarica*, 43–71.
Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology, 25*(4), 463–492. https://doi.org/10.1080/02643290701674851
Kolk, H., & Heeschen, C. (1992). Agrammatism, paragrammatism and the management of language. *Language and Cognitive Processes, 7*(2), 89–129. https://doi.org/10.1080/01690969208409381
Kristinsson, S., Basilakos, A., Den Ouden, D. B., Cassarly, C., Spell, L. A., Bonilha, L., Rorden, C., Hillis, A. E., Hickok, G., Johnson, L., Busby, N., Walker, G. M., McLain, A., & Fridriksson, J. (2023). Predicting outcomes of language rehabilitation: Prognostic factors for immediate and long-term outcomes after aphasia therapy. *Journal of Speech, Language, and Hearing Research, 66*(3), 1068–1084. https://doi.org/10.1044/2022_JSLHR-22-00347
Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience, 18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150
Lambon Ralph, M. A., Patterson, K., & Plaut, D. C. (2011). Finite case series or infinite single case studies? Comments on "Case series investigations in cognitive neuropsychology" by Schwartz and Dell (2010). *Cognitive Neuropsychology, 28*(7), 466–474. https://doi.org/10.1080/02643294.2012.671765
Lampe, L. F., Hameau, S., Fieder, N., & Nickels, L. (2021). Effects of semantic variables on word production in aphasia. *Cortex, 141*, 363–402. https://doi.org/10.1016/j.cortex.2021.02.020
Landrigan, J.-F., Zhang, F., & Mirman, D. (2021). A data-driven approach to post-stroke aphasia classification and lesion-based prediction. *Brain, 144*(5), 1372–1383. https://doi.org/10.1093/brain/awab010
Lee, J., & Thompson, C. K. (2015). Phonological facilitation effects on naming latencies and viewing times during noun and verb naming in agrammatic and anomic aphasia. *Aphasiology, 29*(10), 1164–1188. https://doi.org/10.1080/02687038.2015.1035225
Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776
Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.
Lorca-Puls, D. L., Gajardo-Vidal, A., Mandelli, M. L., Illán-Gala, I., Ezzes, Z., Wauters, L. D., Battistella, G., Bogley, R., Ratnasiri, B., & Licata, A. E. (2023). Neural basis of speech and grammar symptoms in non-fluent variant primary progressive aphasia spectrum. *Brain, 147*(2), 607–626. https://doi.org/10.1093/brain/awad327
Lord, F. M. (1980). *Application of item response theory to practical testing problems*. Erlbaum.
Luzzatti, C., Raggi, R., Zonca, G., Pistarini, C., Contardi, A., & Pinna, G.-D. (2002). Verb–noun double dissociation in aphasic lexical impairments: The role of word frequency and imageability. *Brain and Language, 81*(1-3), 432–444. https://doi.org/10.1006/brln.2001.2536
MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286–1307. https://doi.org/10.1080/02687038.2011.589893
Malyutina, S., Richardson, J. D., & den Ouden, D. (2016). Verb argument structure in narrative speech: Mining AphasiaBank. *Seminars in Speech and Language, 37*(01), 034–047. https://doi.org/10.1055/s-0036-1572383
Matchin, W., Basilakos, A., Stark, B. C., den Ouden, D.-B., Fridriksson, J., & Hickok, G. (2020). Agrammatism and paragrammatism: A cortical double dissociation revealed by lesion symptom mapping. *Neurobiology of Language, 1*(2), 208–225. https://doi.org/10.1162/nol_a_00010
Matchin, W., & Hickok, G. (2020). The cortical organization of syntax. *Cerebral Cortex, 30*(3), 1481–1498. https://doi.org/10.1093/cercor/bhz180
Mauner, G. (2015). The representation and processing of participant role information. In R. G. De Almeida & C. Manouilidou (Eds.), *Cognitive science perspectives on verb representation and processing* (pp. 77–99). Springer International. https://doi.org/10.1007/978-3-319-10112-5_4
Mätzig, S., Druks, J., Masterson, J., & Vigliocco, G. (2009). Noun and verb differences in picture naming: Past studies and new evidence. *Cortex, 45*(6), 738–758. https://doi.org/10.1016/j.cortex.2008.10.003
McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition, 33*(7), 1174–1184. https://doi.org/10.3758/BF03193221

Meulders, M., & Xie, Y. (2004). Person-by-item predictors. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models* (pp. 213–240). Springer. https://doi.org/10.1007/978-1-4757-3990-9_7

Miceli, G., Silveri, M. C., Nocentini, U., & Caramazza, A. (1988). Patterns of dissociation in comprehension and production of nouns and verbs. *Aphasiology, 2*(3–4), 351–358. https://doi.org/10.1080/02687038808248937

Miceli, G., Silveri, M. C., Villa, G., & Caramazza, A. (1984). On the basis for the agrammatic's difficulty in producing main verbs. *Cortex, 20*(2), 207–220. https://doi.org/10.1016/S0010-9452(84)80038-6

Mitchum, C. C., Ritgert, B. A., Sandson, J., & Berndt, R. S. (1990). The use of response analysis in confrontation naming. *Aphasiology, 4*(3), 261–279. https://doi.org/10.1080/02687039008249079

Mohr, J. P. (1976). Broca's area and Broca's aphasia. In H. Whitaker & H. Whitaker (Eds.), *Studies in neurolinguistics* (Vol. 1, pp. 201–233). Academic Press.

Mohr, J. P., Pessin, M. S., Finkelstein, S., Funkenstein, H. H., Duncan, G. W., & Davis, K. R. (1978). Broca aphasia: Pathologic and clinical. *Neurology, 28*(4), 311–324. https://doi.org/10.1212/WNL.28.4.311

Nickels, L. (1995). Getting it right? Using aphasic naming errors to evaluate theoretical models of spoken word recognition. *Language and Cognitive Processes, 10*(1), 13–45. https://doi.org/10.1080/01690969508407086

Nickels, L., Lampe, L. F., Mason, C., & Hameau, S. (2022). Investigating the influence of semantic factors on word retrieval: Reservations, results and recommendations. *Cognitive Neuropsychology, 39*(3–4), 113–154. https://doi.org/10.1080/02643294.2022.2109958

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1, Pt.2), 1–25. https://doi.org/10.1037/h0025327

Park, H., Obermeyer, J., Kornisch, M., Hall, J., & Ontario, C. (2023). Semantic aspects of verb production in various discourse tasks in people with nonfluent aphasia. *American Journal of Speech-Language Pathology, 32*(5S), 2418–2429. https://doi.org/10.1044/2023_AJSLP-22-00293

Perret, C., & Bonin, P. (2019). Which variables should be controlled for to investigate picture naming in adults? A Bayesian meta-analysis *Behavior Research Methods, 51*(6), 2533–2545. https://doi.org/10.3758/s13428-018-1100-1

Pinker, S. (2013). *Learnability and cognition, new edition: The acquisition of argument structure*. The MIT Press. https://doi.org/10.7551/mitpress/9700.001.0001

Rezaii, N., Mahowald, K., Ryskin, R., Dickerson, B., & Gibson, E. (2022). A syntax–lexicon trade-off in language production. *Proceedings of the National Academy of Sciences, 119*(25), Article e2120203119. https://doi.org/10.1073/pnas.2120203119

Rezaii, N., Ren, B., Quimby, M., Hochberg, D., & Dickerson, B. C. (2023). Less is more in language production: An information-theoretic analysis of agrammatism in primary progressive aphasia. *Brain Communications, 5*(3), Article fcad136. https://doi.org/10.1093/braincomms/fcad136

Rofes, A., Capasso, R., & Miceli, G. (2015). Verb production tasks in the measurement of communicative abilities in aphasia. *Journal of Clinical and Experimental Neuropsychology, 37*(5), 483–502. https://doi.org/10.1080/13803395.2015.1025709

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and Language, 37*(3), 440–479. https://doi.org/10.1016/0093-934X(89)90030-8

Saffran, E. M., Schwartz, M. F., & Marin, O. S. (1980). The word order problem in agrammatism: II. Production. *Brain and Language, 10*(2), 263–280. https://doi.org/10.1016/0093-934X(80)90056-5

Schwartz, M. F., & Brecher, A. (2000). A model-driven analysis of severity, response characteristics, and partial recovery in aphasics' picture naming. *Brain and Language, 73*(1), 62–91. https://doi.org/10.1006/brln.2000.2310

Schwartz, M. F., & Dell, G. S. (2010). Case series investigations in cognitive neuropsychology. *Cognitive Neuropsychology, 27*(6), 477–494. https://doi.org/10.1080/02643294.2011.574111

Schwartz, M. F., Dell, G. S., Martin, N., Gahl, S., & Sobel, P. (2006). A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language, 54*(2), 228–264. https://doi.org/10.1016/j.jml.2005.10.001

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Shapiro, L. P., Zurif, E., & Grimshaw, J. (1987). Sentence processing and the mental representation of verbs. *Cognition, 27*(3), 219–246. https://doi.org/10.1016/S0010-0277(87)80010-0

Thompson, C. K. (2003). Unaccusative verb production in agrammatic aphasia: The argument structure complexity hypothesis. *Journal of Neurolinguistics, 16*(2–3), 151–167. https://doi.org/10.1016/S0911-6044(02)00014-3

Thompson, C. K., Lange, K. L., Schneider, S. L., & Shapiro, L. P. (1997). Agrammatic and non-brain-damaged subjects' verb and verb argument structure production. *Aphasiology, 11*(4–5), 473–490. https://doi.org/10.1080/02687039708248485

Thompson, C. K., Lukic, S., King, M. C., Mesulam, M. M., & Weintraub, S. (2012). Verb and noun deficits in stroke-induced and primary progressive aphasia: The Northwestern Naming Battery. *Aphasiology, 26*(5), 632–655. https://doi.org/10.1080/02687038.2012.676852

Thompson, C. K., Shapiro, L. P., Li, L., & Schendel, L. (1995). Analysis of verbs and verb argument structure: A method for quantification of aphasic language production. *Clinical Aphasiology, 23,* 121–140.

Thorne, J., & Faroqi-Shah, Y. (2016). Verb production in aphasia: Testing the division of labor between syntax and semantics. *Seminars in Speech and Language, 37*(01), 023–033. https://doi.org/10.1055/s-0036-1571356

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Lawrence Erlbaum.

Vermeulen, J., Bastiaanse, R., & Van Wageningen, B. (1989). Spontaneous speech in aphasia: A correlational study. *Brain and Language, 36*(2), 252–274. https://doi.org/10.1016/0093-934X(89)90064-3

Vigliocco, G., Vinson, D. P., Druks, J., Barber, H., & Cappa, S. F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neuroscience & Biobehavioral Reviews, 35*(3), 407–426. https://doi.org/10.1016/j.neubiorev.2010.04.007

Walker, G. M., & Hickok, G. (2016). Bridging computational approaches to speech production: The semantic–lexical–auditory–motor model (SLAM). *Psychonomic Bulletin & Review, 23*(2), 339–352. https://doi.org/10.3758/s13423-015-0903-7

Walker, G. M., Hickok, G., & Fridriksson, J. (2018). A cognitive psychometric model for assessment of picture naming abilities in aphasia. *Psychological Assessment, 30*(6), 809–826. https://doi.org/10.1037/pas0000529

Walker, G. M., & Schwartz, M. F. (2012). Short-form Philadelphia Naming Test: Rationale and empirical evaluation. *American Journal of Speech-Language Pathology, 21*(2), S140–S153. https://doi.org/10.1044/1058-0360(2012/11-0089)

Whitworth, A., Webster, J., & Howard, D. (2013). *A cognitive neuropsychological approach to assessment and intervention in aphasia: A clinician's guide* (2nd ed.). The Psychology Press.

Wilson, S. M., Bautista, A., & McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *NeuroImage, 171,* 62–74. https://doi.org/10.1016/j.neuroimage.2017.12.068

Wilson, S. M., Brandt, T. H., Henry, M. L., Babiak, M., Ogar, J. M., Salli, C., Wilson, L., Peralta, K., Miller, B. L., & Gorno-Tempini, M. L. (2014). Inflectional morphology in primary progressive aphasia: An elicited production study. *Brain and Language, 136,* 58–68. https://doi.org/10.1016/j.bandl.2014.07.001

Wilson, S. M., Entrup, J. L., Schneck, S. M., Onuscheck, C. F., Levy, D. F., Rahman, M., Willey, E., Casilio, M., Yen, M., Brito, A. C., Kam, W., Davis, L. T., de Riesthal, M., & Kirshner, H. S. (2023). Recovery from aphasia in the first year after stroke. *Brain, 146*(3), 1021–1039. https://doi.org/10.1093/brain/awac129

Wilson, S. M., Isenberg, A. L., & Hickok, G. (2009). Neural correlates of word production stages delineated by parametric modulation of psycholinguistic variables. *Human Brain Mapping, 30*(11), 3596–3608. https://doi.org/10.1002/hbm.20782

Zingeser, L. B., & Berndt, R. S. (1990). Retrieval of nouns and verbs in agrammatism and anomia. *Brain and Language, 39*(1), 14–32. https://doi.org/10.1016/0093-934X(90)90002-X